DOCUMENT RESUME

ED 090 555 CS 201 216

AUTHOR Veal, L. Ramon
TITLE Syntactic Measures and Rated Quality in the Writing of Young Children. Studies in Language Education, Report No. 8.
INSTITUTION Georgia Univ., Athens. Dept. of Language Education.
PUB DATE Feb 74
NOTE 16p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Composition (Literary); Educational Research; Elementary Education; *Language Arts; *Sentence Structure; *Syntax; *Writing Skills

ABSTRACT

This preliminary investigation grew out of the Project English study of written composition at the University of Georgia from 1963 through 1968. Toward the end of that project, attention was given to possible relationships between the quality of written composition and certain syntactic measures in student writing produced in schools using the project-developed materials. In summary, the syntactic measures studies clearly distinguish between high and low quality writing in the second, fourth, and sixth grades. A general implication of these findings is that the teaching of structural options to enhance maturity in writing might also, at the elementary level, enhance quality. (TO)

SYNTACTIC MEASURES AND RATED QUALITY IN THE WRITING OF YOUNG CHILDREN

by

L. Ramon Veal

Studies in Language Education, Report No. 8
Department of Language Education, The University of Georgia
Athens, Georgia

February, 1974

Introduction. The focus of this preliminary investigation grew out of a larger study of the written composition of young children, namely, the "Project English" study of written composition at the University of Georgia, (1963 to 1968). Specifically, the Georgia English Curriculum Study Center developed materials for an elementary school curriculum in written composition. Part of that curriculum project included the development of an instrument to assess globally the quality of writing produced in schools using the Study Center materials. Toward the end of the project and as data reflecting quality began to come in, attention was divided to possible relationships between levels of rated quality and certain syntactic measures. This report, therefore, is a summary of those preliminary investigations.

As several researchers (Loban, 1963; Hunt, 1965; O'Donnell, 1967) have observed, syntactic measures indicate maturity in writing. Hunt, who has used the minimal terminal syntactic unit[1] as a basic unit of measurement in syntactic analysis, has noted significant differences between grades four, eight, and twelve. O'Donnell found differences between grades three, five, and seven. In each case, however, the measurement involved structural or syntactic complexity and not estimated quality.

Mellon (1967), Bateman and Zidonis (1966), O'Hare (1973) and many other researchers have employed syntactic measures as criteria in experimental studies of written composition. Except for Mellon's and O'hare's, however, these and similar investigations have not included overall estimates of quality. Moreover, they have concentrated on the sentence as the unit of measurement and analysis. And even though they have demonstrated that a more felicitous

---

[1]Generally referred to as T-unit, it is described as one main clause and its modifiers--the shortest grammatically allowable unit into which an essay can be segmented without leaving residue.

-2-

use of a greater variety of structures can be taught, they have not shown that a direct connection exists between syntactic complexity and overall quality. The one sub-sample that Mellon checked for both quality and sentence structure revealed no statistically significant correlations between the two variables of syntax and quality. O'Hare's experimental group, however, the one that demonstrated greater syntactic maturity, also produced writing that was judged to be significantly better than that of the control group, the one with normal syntactic maturity. Still, specific aspects of syntax were not compared with overall quality.

In short, syntactic measures are objective, and they do reflect maturity in writing, maybe even quality. It was therefore the purpose of this study to compare observed differences in the estimated quality of writing and differences in syntactic complexity. Furthermore, it was assumed that the degree to which a syntactic measure correlated with global ratings would be one index to the validity of the developed instrument.

Test Development. First, a brief description of the developed instrument is in order. In the view of several researchers (Braddock, 1963; Findley, 1963; Derrick, 1964), evaluation of children's writing should focus upon the actual composition itself--the child's written product. One possibility for such evaluation is the product-scale class of instrument, an example of which is the S.T.E.P. Essay Test. It evaluates the "product" (that is, the student's written composition) by comparing it with other "products" to derive a relative measure of its merit. In the development of a product-scale composition instrument, products which are actual samples of student writing must be selected to serve as models. In order to provide the necessary controls in developing these models, it is important that the models reflect common criteria, that they be selected by trained raters, that they come from several samples; and, above all,

that these models can then be used by minimally trained raters when the writing samples to be evaluated are produced under standardized examination situations paralleling those of the models.

Essentially, the test-development task was to construct a product-scale instrument that would yield reliable estimates of the overall quality of writing samples produced by elementary school children. Thus, it was necessary to begin by collecting samples of pupil writing from which model or comparison essays could be selected. Steps in this process included:

1) reviewing the literature to determine which criteria were typically used in evaluating written composition at the elementary level (33 were found);
2) identifying from this list those criteria that were to be used in selecting the model essays--nine criteria were selected by 32 cooperating teachers who ranked the criteria as to importance;
3) obtaining under standardized conditions writing samples from classes cooperating with the English Curriculum Study Center;
4) rating these papers, according to the selected criteria, using trained raters on a seven-point scale;
5) selecting from these rated papers reliably rated (.87) comparison essays that represent--according to the criteria--high, average, and low quality (point six, four, and two) on the rating scale;
6) obtaining additional writing samples on each of the eight different topics to use in a reliability check; and, finally,

7) rating these additional papers, using both trained and untrained raters as well as the selected comparison essays for points of reference on the rating scale.

Rater reliability data were collected from several groups of raters on several separate occasions over more than two years. Included in the estimate of rater reliability were trained raters and raters with only a minimal amount of instruction. Data were collected to show the reliability of one rater, the reliability among different numbers of raters, inter-reliability of topics or forms, and the reliability of both trained and untrained raters.

Inter-rater reliability was found to be high (.70 or better). In fact, it was significantly higher than that achieved in typical ratings made by English teachers (.50) and comparable to that typically reported (.70) for trained raters (Diederich, 1964). Test-retest reliability coefficients for the different forms ranged from .58 to .89, and a single test-retest check of reliability on the same form proved to be .71.

Validity was assessed by comparing ratings made by the criteria method with those made by the comparison method. Three trained raters rated the essays by both methods. For the criteria method, reliability coefficients with a range of .49 to .86 and a mean of .66 were obtained. For the comparison method with the same raters and the same papers, coefficients with a range of .52 to .80 and a mean of .64 were obtained. Differences between inter-rater reliability for the two methods were not statistically significant. Both methods produced reliable ratings, and overall agreement between the two methods (.79) provides at least one estimate of validity.

For each form, the ratings by comparison were higher than the ratings by criteria; and for three forms the differences were significant. Both methods led to essentially the same conclusions with regard to a ranking of the essays, but the two methods led to different conclusions about the absolute quality of the papers.

On this particular test of validity, the developed instrument provided a valid index to the relative quality of the sample essays but did not yield estimates of particular levels of quality (1-7) comparable to estimates based on criteria.

The syntax study, it should be noted, was based on the relative quality--and not the absolute quality--of the papers.

Sample for Syntax Study. In another school (that is, one not cooperating with the ECSC), eight forms of the developed instrument were administered to all second, fourth, and sixth grade children: four forms in the fall of 1967 and four in the spring of 1968. Twenty-seven children were selected, in random fashion, from each of the three grades. Thus, eight papers from 81 children (648 papers in all) were available for rating and analysis. All papers were rated by three raters and an average taken. These averages were then summed across all eight papers to provide a single estimate of quality. Finally, these estimates of quality were ranked and divided into three equal groups (high, middle, and low) for each of the three grades. The selected sample, therefore, had equal-size groups of children in each of three levels (H, M, L) for each of three grades (6, 4, 2). Table I shows this grouping.

TABLE I

NUMBERS OF CHILDREN BY LEVEL AND GRADE

Rated Quality Level	Grade 2	4	6	
Low	9	9	9	
Middle	9	9	9	
High	9	9	9	
	27	27	27	81

-6-

Additional data on the children showed the grade groupings not significantly different on either mental or reading ability. Also, the range for each grade followed a normal distribution and range in both I.Q. scores (mean of 104) and grade placement reading scores (means at or above grade level). The possible relationships between I.Q. scores, reading scores, rater estimates of quality, and the several syntactic measures constitute another study that we have projected to follow this preliminary investigation into quality and syntax. These subject variables, since they did not differentiate groups, are not considered in the present study of syntax.

It should also be noted that a composite of eight papers, as used in this study, provides as reliable an estimate of rated quality as one is likely to get (Diederich, 1964). Also, the syntactic data are comparable to those of O'Donnell's (1967) similar sample. In short, these procedures provided for a cross section of grade-level and rated-quality writing of elementary age children, writing that could then be looked at syntactically.

Procedures for Syntax Study. First of all, each paper was segmented into T-units. Frequency counts, based on a composite of all eight papers, were then made of the following:[2]

1) total words
2) total T-units
3) words per T-unit
4) subordinate clauses
5) total clauses
6) clauses per T-unit
7) coordinators of T-units

---

[2]Acknowledgment is here given to Dan Ward who, as part of his thesis study at the University of Georgia made these counts.

These counts are reported in Table 2.

For the sake of comparison, those counts were translated to a base of 100 T-units and are reported in Table 3. One further step involved a breakdown of the subordinate clauses into nominal, adverbial, and adjective clauses. This breakdown is shown in Table 4.

Differences between levels and grades were checked for statistical significance and are reported in Tables 5 and 6.

Results of the Syntactic Study. As several researchers (Biesbrock, 1968; Cartwright, 1968; and Martin, 1968) have shown, when evaluation of written composition is based on a timed sample, length of composition is likely to be a factor in rated quality. In this study, significant differences in composition length (total words) were found between all grades and between most levels of quality. Generally speaking, the same situation was found to be true of the total number of T-units and clauses, except at the thresholds between grades--that is, between high second and low fourth grades and between high fourth and low sixth grades (see Table 2). Also, as both Hunt (1965) and O'Donnell (1967) have shown, T-unit and clause length are significant indices of maturity--that is, they correlate highly with grade level.

In this study, too, clause length, T-unit length, and number of clauses per T-unit steadily increased (at a statistically significant level) from grade to grade. In fact, the data correspond almost precisely with O'Donnell's (1967) cross-sectional sample. Thus, the writing specimens under consideration were reliably rated as to global quality, and they also reflected the usual distinctions in structural complexity. The questions, then, of this preliminary study centered around possible connections between reliable indices of structural complexity and reliable estimates of quality.

Table 5 summarizes syntactic distinctions between levels of quality within grades. It shows that T-unit length was related to differing quality levels at each of the three grades. It distinguished differently, however, for each of the grades--at grade two, between low and middle and between low and high but not between middle and high. At the fourth grade it distinguished only between low and high levels; and, at the sixth grade, it distinguished between middle and high and between low and high but not between low and middle. For all grades, though, an increase in at least one word in T-unit length differentiated between low and high levels of quality.

An increase in the number of subordinate clauses clearly distinguished between all levels of quality except between low and middle second grade groups. Apparently, this increase in subordinate clauses accounted for much of the increase in T-unit length, for, as Table 5 shows, clause length differentiated only between low and high second grades. Also, as Tables 4 and 5 indicate, the increased number of subordinate clauses included all kinds of clauses at all levels and grades; but adverbials increased more than did either nominal or adjective clauses, both of which remained rather stable for the second and fourth grades, increasing mainly at grade six.

The nominal clauses, as Table 4 suggests, included very little dialog except at the sixth grade, high level. In addition, the positions of the adverbial clauses showed a balance increase through all grades and levels, with the medial position exhibiting the highest percentage increase.

Hunt (1965) observed that the use of simple coordinators of T-units, like and and but, are marks of immature writing, and Potter (1967) noted the same characteristic in identifying the structures of "good" and "poor" writing at the tenth grade. The absence of such coordinators in this sample clearly marked the writing of high fourth graders and all levels of quality at the sixth grade.

Table 6 lists significant distinctions between levels of quality across grades. As the table indicates, words per T-unit, words per clause, and clauses per T-unit distinguish between almost all levels of quality, with words per T-unit significantly making 22 of 27 possible distinctions in quality. Words per clause make 20 such distinctions, while clauses per T-unit made 17. Thus, T-unit length proved to be the most effective syntactic marker of quality, although two related factors, clause length and number of clauses, also proved effective.

These structural characteristics of sentences distinguished more levels of quality at the second and fourth grades than at the sixth. In fact, the syntax of low quality sixth grade writing was shown to be not essentially different from all levels of the fourth grade. In the breakdown of clauses, in particular, as much variation was found within grades as between them.

Conclusions and Implications. In summary, the syntactic measures studies clearly distinguished between high and low quality writing in the second, fourth, and sixth grades. Also, the addition of subordinate clauses seemed to account more for differing levels of quality within grades than for differences in quality or syntax between grades. Moreover, no one kind of clause seemed to account for qualitative differences more than any other; and, for adverbials, position did not appear to distinguish quality. In short, the same syntactic measure (T-unit length) that has recently been shown to identify maturity in writing appears also to distinguish at least two--and, in some cases, three--levels of quality in these elementary grades.

A general implication of these findings is that the teaching of structural options to enhance maturity in writing might also, at the elementary level, enhance quality.

REFERENCES

Bateman, Donald R. and Zidonis, Frank. THE EFFECT OF A STUDY OF TRANSFORMATIONAL GRAMMAR ON THE WRITING OF NINTH AND TENTH GRADERS. Champaign, Illinois: NCTE, 1966.

Biesbrock, Edieann. THE DEVELOPMENT AND USE OF A STANDARDIZED INSTRUMENT FOR MEASURING COMPOSITION ABILITY IN YOUNG CHILDREN (Grades Two and Three), dissertation completed at the University of Georgia, 1968.

Braddock, Richard, et al. RESEARCH IN WRITTEN COMPOSITION. Champaign, Illinois: NCTE, 1963.

Cartwright, G. Phillip. "Discriminating Power of Several Objective Measures of Writing Ability," paper read at AERA, 1968.

Derrick, Clarence. "Tests of Writing," English Journal. Vol. 53 (1964).

Diederich, P. B. Problems and possibilities of research in the teaching of written composition, in RESEARCH DESIGN AND THE TEACHING OF ENGLISH. Champaign, Illinois: NCTE, 1964.

Findley, Warren. "Improvement in Writing Ability." 20TH YEARBOOK, NCME, 1963.

Hunt, Kellog W. GRAMMATICAL STRUCTURES WRITTEN AT THREE GRADE LEVELS. Champaign, Illinois: NCTE, 1965.

Loban, Walter. THE LANGUAGE OF ELEMENTARY SCHOOL CHILDREN. Champaign, Illinois: NCTE, 1963.

Martin, H. Glenn. "Could Computers Grade Pupil Themes?" S. C. EDUCATION REPORT (Fall, 1968).

Mellon, John C. TRANSFORMATIONAL SENTENCE-COMBINING. Cambridge, Massachusetts: CRP No. 5-8418, Harvard R & D Center Report 1, 1967.

O'Donnell, Roy C., et al. SYNTAX OF KINDERGARTEN AND ELEMENTARY SCHOOL CHILDREN: A TRANSFORMATIONAL ANALYSIS. Champaign, Illinois: NCTE, 1967.

O'Donnell, Roy C. "An Objective Measure of Structural Complexity in Children's Writing," paper read at AERA, 1968.

O'Hare, Frank. SENTENCE COMBINING. Urbana: NCTE, 1973.

Potter, Robert R. "Sentence Structure and Prose Quality." RTE, Vol. 1 (Spring, 1967).

TABLE 2

MEANS OF THREE LEVELS OF QUALITY FOR
THREE GRADES ON SEVERAL SYNTACTIC MEASURES
(Based on Sight Writing Samples)

Grades	Levels of Quality	(M)	Total Words	Total T-Units	Words T-Unit	Subordinate Clauses	Total Clauses	Words/ Clauses	Clauses/ T-Unit	Coordinators
2	Low	(9)	360	62	5.81	7	69	5.22	1.11	11
	Middle	(9)	458	67	6.84	9	76	6.03	1.13	17
	High	(9)	578	75	7.71	17	92	6.28	1.23	14
4	Low	(9)	608	74	8.22	13	87	6.99	1.18	19
	Middle	(9)	808	94	8.60	22	116	6.97	1.23	24
	High	(9)	886	95	9.33	29	124	7.15	1.29	15
6	Low	(9)	729	83	8.78	16	99	7.36	1.19	11
	Middle	(9)	997	106	9.23	35	143	6.97	1.31	18
	High	(9)	1,284	123	10.44	44	167	7.69	1.36	19

TABLE 3

MEANS OF THREE LEVELS OF QUALITY FOR
THREE GRADES ON SEVERAL SYNTACTIC MEASURES
(Based on 100 T-Units)

Grades	Levels of Quality	Total Words	Words/ T-Unit	Subordinate Clauses	Total Clauses	Words/ Clauses	Clauses/ T-Unit	Coordinators
2	Low	581	5.81	11	111	5.22	1.11	18
	Middle	684	6.84	13	113	6.03	1.13	25
	High	771	7.71	23	123	6.28	1.23	19
4	Low	822	8.22	18	118	6.99	1.18	26
	Middle	860	8.60	23	123	6.97	1.23	26
	High	933	9.33	29	129	7.15	1.29	16
6	Low	878	8.78	19	119	7.36	1.19	13
	Middle	923	9.23	31	131	6.97	1.31	17
	High	1044	10.44	36	136	7.69	1.36	15

TABLE 4

SUBORDINATE CLAUSES PER 100 T-Units

Grades	Levels of Quality	N	Nominal		Adverbial			Adjective
			Dialog	Non-Dialog	Initial	Medial	Final	
2	Low	9	.18	5.19	1.97	.35	2.69	1.24
	Middle	9	.00	3.48	2.64	.66	6.13	.66
	High	9	.59	5.63	5.33	1.04	5.66	4.29
4	Low	9	.15	4.05	3.91	1.35	6.61	1.65
	Middle	9	.23	4.02	7.91	.83	6.98	3.54
	High	9	.46	6.55	8.19	2.11	9.36	3.86
6	Low	9	.00	3.88	5.22	1.34	6.42	2.94
	Middle	9	.31	7.81	5.15	1.95	8.95	7.93
	High	9	1.17	6.86	8.94	2.35	9.67	6.86

TABLE 5

DIFFERENCES BETWEEN RATED LEVELS OF QUALITY ON SELECTED SYNTACTIC MEASURES FOR THREE GRADES

(Based on 100 T-Units)

Differences Between Levels (By Grades)	Words per T-Unit	Clauses per T-Unit	Words per Clause	Subordinate Clauses			
				Nominal	Adverbial	Adjective	Coordinators
Grade Two							
Low & Middle	1.03*	.02	.81	-2*	4*	0	7
Middle & High	.87	.10**	.25	3**	3*	3**	-6
Low & High	1.90**	.12**	1.06*	1	7**	3**	1
Grade Four							
Low & Middle	.38	.05*	-.02	0	4*	2*	0
Middle & High	.73	.06*	.18	3*	3	0	-10*
Low & High	1.11*	.11**	.6	3*	7**	2*	-10*
Grade Six							
Low & Middle	.45	.12**	-.39	4*	3	5**	4
Middle & High	1.21*	.05*	.72	0	5*	-1	-2
Low & High	1.66*	.17**	.33	4*	8**	4**	2

* Difference Significant at .05 level.

** Difference Significant at .01 level.

TABLE 6

SIGNIFICANT* DIFFERENCES BETWEEN RATED LEVELS OF QUALITY ON SELECTED SYNTACTIC MEASURES ACROSS THREE GRADES

Grade and Level Differences	Words per T-Unit	Clauses per T-Unit	Words per Clause	Subordinate Clauses			Coordinators
				Nom.	Adv.	Adj.	
L2-L4	x	x	x	-	x	-	-
L2-M4	x	x	x	-	x	x	-
L2 H4	x	x	x	-	x	x	-
L2-L6	x	x	x	-	x	-	-
L2-M6	x	x	x	x	x	x	-
L2-H6	x	x	x	x	x	x	-
M2-L4	x	-	x	-	-	-	-
M2-M4	x	x	x	-	x	x	-
M2-H4	x	x	x	x	x	x	-
M2-L6	x	-	x	-	-	-	-
M2-M6	x	x	x	x	x	x	-
M2-H6	x	x	x	x	x	x	-
H2-L4	-	-	x	-	-	-	-
H2-M4	x	-	x	-	-	-	-
H2-H4	x	-	x	-	-	-	-
H2-L6	x	-	x	-	-	-	-
H2-M6	x	x	x	-	-	x	-
H2-H6	x	x	x	-	x	x	-
L4-L6	-	-	-	-	-	-	-x
L4-M6	x	x	-	x	-	x	-
L4-H6	x	x	x	x	x	x	-x
M4-L6	-	-	-	-	-	-	-x
M4-M6	x	x	-	x	-	x	-
M4-H6	x	x	x	x	-	x	-x
H4-L6	-	-x	-	-	-	-	-
H4-M6	-	-	-	-	-	x	-
H4-H6	x	-	-	-	-	x	-

* (x) indicates observed difference is significant (at least at the .05 level).